11-1-1996

SHAPE RECONSTRUCTION BY A BINOCULAR FIXATING SYSTEM

Moses Chan
Purdue University School of Electrical and Computer Engineering

Zygmunt Pizlo
Purdue University School of Electrical and Computer Engineering

David Chelberg
Purdue University School of Electrical and Computer Engineering

Shape Reconstruction by a Binocular Fixating System

Moses Chan
Zygmunt Pizlo
David Chelberg

TR-ECE 96-18
November 1996

School of Electrical
and Computer Engineering
Purdue University
West Lafayette, Indiana 47907-1285

SHAPE RECONSTRUCTION BY A BINOCULAR FIXATING SYSTEM

Moses Chan

Zygmunt Pizlo

David Chelberg

School of Electrical and Computer Engineering

Purdue University

West Lafayette, IN 47907

This work is supported in part by the NIH grant # 1 R01 CA62243-01A1

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

In this paper, we present an algorithm which is a possible model of the human binocular shape reconstruction. In this new model, the shape of an object can be reconstructed from two perspective views with as few as three points, without explicitly estimating the object's pose. Simulation experiments show that the algorithm provides reliable results in the presence of noise. The psychological plausibility of this algorithm has been tested in a psychophysical experiment on shape constancy. Results of this experiment contradict prior models of human shape perception and are consistent with some (but not all) properties of our new algorithm.

1. INTRODUCTION

In this paper, we consider shape perception in the case of both human vision and computer vision. Since it is known that human observers are extremely efficient in shape perception tasks, we believe that a psychologically plausible algorithm for shape recognition and reconstruction is likely to be computationally efficient, and thus lead to progress in computer vision applications.

Shape is defined conventionally as the property of the contour of a figure or of the surface of an object that is invariant under rigid motion (translation, rotation) and size scaling (i.e. under the group of similarity transformations). Thus, shape does not change if angles or ratios of distances do not change. This definition implies that shapes are characteristic properties of rigid objects. Therefore, shapes can be used to recognize objects.

Research on shape perception was initiated by psychologists at the beginning of this century (e.g. Koffka, 1935 [Kof35]). Psychologists have been primarily concerned with a phenomenon called shape constancy which is closely related to shape recognition. Shape constancy refers to the fact that the percept of the shape of a given object remains constant despite changes in the shape of the object's retinal image. The retinal image may change because of changes in the orientation and distance of the object relative to the observer.

We consider here shape perception in the case of a conventional vision system which obtains visual information about a 3-D scene from the 2-D perspective images. Since perspective images do not provide the 3-D information directly, the percept of shape cannot involve similarity properties computed from the retinal images. Instead,

one has to 1) use invariants of perspective transformation, or 2) reconstruct the 3-D similarity structure. We consider first invariants.

1.11 The role of invariants

An invariant of a group of transformations is a property which remains constant, under these transformations. Since the camera image is a perspective transformation of a scene, the smallest group whose invariants are preserved in the camera image is a projective group. Therefore, if shape constancy (or shape recognition) involves invariants, then these should be projective invariants. This conjecture was first made by Cassirer (1938/1944) [Cas44] , Courant and Robbins (1941) [CR41] and then by Gibson (1950) [Gib50].

If invariants are to be useful in shape recognition, they must be general case invariants. A general case view invariant is a property which can be computed for any 3-D point set. At the same time an invariant should be non-trivial, i.e., there should exist at least two different sets of points such that the value of this invariant would be different. Clearly, if an invariant has the same value for all objects, it is useless because it cannot distinguish different objects.

In the past 50 years, projective invariants have been used as an explanation for shape constancy in human vision [Gib50] [Joh77] [Cut86], and as a tool for solving shape recognition problems in computer vision [DH73] [Wei88] [BBHP92]. However, there are problems with using projective invariants in shape recognition and shape constancy. It is known that non-trivial general case view invariants do not exist for arbitrary 3-D point sets under 3-D to 2-D transformations [BWR90]. The absence of general case view invariants for 3-D shapes has led researchers to use either general case invariants for 2-D shapes or special case invariants for structured 3-D point sets [RFZM93]. In order to obtain general case view invariants for 3-D shapes, it is necessary to use more than one view [KvD91] [BBP92] [BBHP92]. However, if two views are available, it is possible to reconstruct the 3-D shape (i.e. the similarity structure) of an object (see the next section).

In the case of a single view of a 2-D shape there are general case projective invariants. However, Astrom (1995) [Ast95] showed that a given planar curve can be projectively transformed so that it becomes arbitrarily close to any other curve. As a result, even if two curves are projectively different, they may be difficult or even impossible to be distinguished in the presence of (Euclidean) noise in the image. In psychology there are problems with using projective invariants as a model of shape constancy even in the absence of noise. Since all (convex) quadrilaterals are projectively equivalent, it follows that if projective invariants were irivolved in human shape perception, human observers would not be able to discriminate among different quadrilaterals. However, Stavrianos (1945 [Sta45]) and Pizlo [Piz94] showed that human subjects can reliably discriminate among quadrilaterals from a single perspective image. Therefore, projective invariants are not psychologically plausible. Pizlo (1994 [Piz94]) showed that an adequate explanation of shape constancy in the case of planar shapes requires a different type of invariants called quasi- (or model-based) invariants (Pizlo & Rosenfeld, 1992 [PR92]). Quasi-invariants are geometrical properties that are computed from two shapes (rather than from one shape, as in the case of conventional invariants) related by a given transformation. In the case of shape constancy, quasi-invariants involve a shape of an object and a perspective image and they allow one to determine whether this image could halve been produced by this shape. Quasi-invariants also exist in the case of 3-D shapes (see Weinshall, 1993 [Wei93], for quasi-invariant of parallel projection; and Pizlo and Loubier, 1996 [PLed], for quasi-invariant of perspective projection).

To summarize the role of invariants in shape perception. Projective invariants are not likely to be useful in shape recognition (constancy) because: 1) there are no general case invariants in the case of a single view of a 3-D shape; 2) general case projective invariants exist in the case of a single view of a 2-D shape but these invariants are neither psychologically plausible nor computationally efficient in the presence of noise. Shape recognition (constancy) is more likely to involve quasi-invariants and

such invariants exist both for planar and solid shapes. Note, however, that quasi-invariants (or any other invariants) computed from a single image cannot be used in the case of a novel shape because a single image does not allow one to reconstruct the shape. Shape reconstruction is possible from a single or multiple images if depth cues are used. Alternatively, reconstruction may involve quasi-invariants computed from two images. In the next section, we briefly describe prior mathematical solutions for shape reconstruction.

1.2 The role of reconstruction

Most theories that have been formulated in the area of human vision, have assumed that shape reconstruction is preceded by estimating the object's depth (i.e. viewing distance), and the object's 3-D orientation (i.e. the object's pose). Depth and orientation have been, in turn, assumed to be computed from depth cues such as texture, motion, shading, binocular disparity and vergence [Roc83]. However, it is known that depth cues are not a reliable source of information that could be used by human observers in shape perception [TN94] [Joh91]. Therefore, one should consider the possibility of reconstructing an object's shape without estimating its depth and orientation. It is known that such reconstruction is possible if two perspective images are available.

This approach was first used in 1959 by Thompson [Tho59]. He showed that two perspective views of 5 non-coplanar points are sufficient for unique reconstruction of the Euclidean structure of these points (up to size scaling). Thompson's solution involved an iterative method. This solution was subsequently simplified by Longuet-Higgins [LH81] (8-point algorithm) who showed that in the case of 8 non-coplanar points, the Euclidean structure can be computed by solving a set of 8 linear equations. It is important to note that these two methods can only be applied to perspective views. If the range of a visual scene in depth is small as compared to the viewing distance (a situation which practically never happens in everyday life, although is often used in laboratory studies), perspective projection reduces to parallel projection.

It is well known that two parallel projections of a 3-D shape do not uniquely determine the 3-D Euclidean structure of the shape, regardless of the number of distinctive points [AB89] [KvD91]. In order to obtain a unique solution, the observer must use either at least three projections [Ull79] [AB89] [TK90], or two projections plus depth cues. Since we are interested in modeling shape perception by human beings in everyday life situations, we will model the case of binocular viewing without using depth cues (because depth cues are unreliable), and we will concentrate on the case of two perspective views that are not equivalent to parallel projections.

The 8-point algorithm [LH81] was independently developed by Tsai and Huang [TH84] to estimate motion parameters from perspective image sequences. They also proved that the reconstruction of the 8-point algorithm is unique if the points satisfy certain geometrical properties (refer to [TH84] for details). It is irnportant to point out that in the presence of noise the performance of the 8-point algorithm degrades if the object's size is small relative to the viewing distance. As discussed before, under such conditions two perspective projections reduce to two parallel projections, and therefore two views are not sufficient to determine the 3-D Euclidean structure uniquely. However, even if the object's size is not small relative to the viewing distance, the solution of the 8-point algorithm is still extremely sensitive to noise in the images. To cope with the noisy image data, Spetsakis et al. [SA89] provided an iterative mean square error minimization method. Recently, Hartley [Har95] developed a non-iterative method to solve this problem.

To summarize, shape can be reconstructed from two perspective views without using depth cues. Prior algorithms that perform such a reconstruction allowed all degrees of freedom in the relative orientation and position of one camera relative to the other. Since in this paper we investigate a model of human binocular vision, we will use some natural constraints on the relative orientation between the two cameras. Namely, we assume that the two cameras represent a fixating vision system, in which the two visual axes intersect and the cameras do not rotate around their visual axes. Our algorithm requires as few as three corresponding points and performs reliably in

the presence of noise. To our knowledge, it is the simplest modification of the 8-point algorithm thus far.

The rest of this paper is organized as follows. In Section 2, we describe our new algorithm of shape reconstruction for the case of a fixating binocular system. Section 3 presents the results of the simulations. In Section 4, we describe a psychophysical experiment on human shape constancy. Results of this experiment show that human shape reconstruction does not involve depth cues. Instead, it involves images of occluding contours. These results are consistent with some (but not all) aspects of our new algorithm. Section 5 summarizes the main results presented in this paper.

2. COMPUTER ALGORITHM

The geometry of the formation of images in the two cameras is shown in Figure 2. We assume that the center of projection of the right camera is at the origin of the coordinate system and the focal length is equal to one. The object's points (that are to be reconstructed) are represented in this coordinate system. The image points in the right camera are related to the object points by the rules of perspectivity. The left camera is translated and rotated relative to the right camera. We can assume, without restricting the generality, that the left camera coincides with the right camera, and the image in the left camera is obtained after translating and rotating the 3-D object relative to this camera. In other words, instead of using two cameras that view the object from different directions, we can use one camera and take the images for two different orientations and positions of the object. Consider an object point P. We use the following notation:

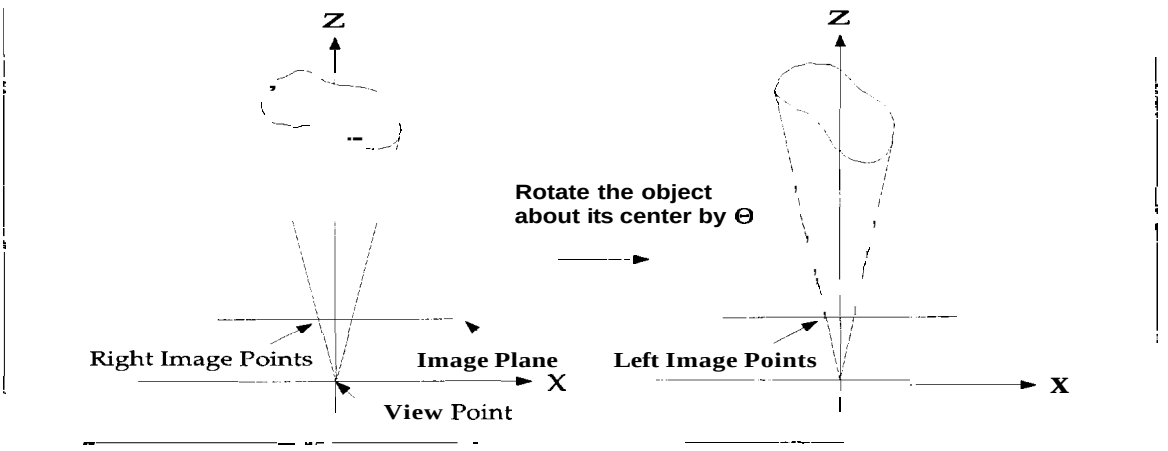

Fig. 2.1. Geometry of the problem.

$(R = R_1, R_2, R_3)$ - an object point P in the first view

($L = L_x, L_y, L_z$) - an object point P in the second view after translation T and rotation **Rot**

$(\mathbf{r} = r_x, r_y, 1)$ - the image of **R**

$(1 = l_x, l_y, 1)$ - the image of L

The two positions of the object point P are represented by the following equation:

$$\mathbf{L} = \mathbf{Rot}(\mathbf{R} + \mathbf{T}) \tag{2.1}$$

We will first follow the method described by Longuet-Higgins (1981) [LH81] in which equation 2.1 is transformed in such a way that it relates the image points in the first view with the image points in the second view by a linear transformation with 8 free parameters. When such a relation is obtained for 8 corresponding points, the problem reduces to solving 8 simultaneous linear equations with eight unknowns. Then, we will incorporate constraints inherent in the fixating binocular system. This will reduce the number of unknowns to three. Finally, we will show how the object points can be reconstructed.

2.1 Longuet-Higgins's (1981) derivation

Equation 2.1 can be transformed as follows:

$$\begin{aligned} \mathbf{Rot}^t\mathbf{L} &= \mathbf{R} + \mathbf{T} \\ \mathbf{T} \times (\mathbf{Rot}^t\mathbf{L}) &= \mathbf{T} \times \mathbf{R} \\ \mathbf{R}^t[\mathbf{T} \times (\mathbf{Rot}^t\mathbf{L})] &= 0 \end{aligned} \tag{2.2}$$

The perspective image $(r_x, r_y, 1)$ of an object point (L_x, L_y, L_z) in the camera whose geometry has been defined above is computed as:

$$r_x = R_x/R_z \qquad r_y = R_y/R_z$$

Similarly, for the second view:

$$l_x = L_x/L_z \qquad l_y = L_y/L_z$$

After dividing both sides of equation 2.2 by $(R_z L_z)$ and transforming the cross product of two vectors into a product of two matrices (by introducing a skew-symmetric matrix), we obtain:

$$\begin{bmatrix} r_x & r_y & 1 \end{bmatrix} \begin{bmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{bmatrix} \mathbf{Rot'} \begin{bmatrix} l_x \\ l_y \\ 1 \end{bmatrix} = 0 \qquad (2.3)$$

Equation 2.3 is equivalent to equation 12 in Longuet-Higgins (1981).

2.2 Natural constraints of a binocular fixating observer

The rotation **Rot** and translation **T** in equation 2.1 represent any rigid motion in 3-11, which means that there is no constraint on the relative orientation and position between the two cameras. However, in the case of the human visual system, when the eyes of a human observer are fixated on a particular point of an object, the optical axes of the two eyes are coplanar (approximately). In other words, the elevations of the two eyes are the same. Moreover, each eyeball of the human observer performs rotations with only two degrees of freedom, rather than three. Namely, the rotation around the visual axis (torsion) is not a free parameter in the movements of the eye. Specifically, the magnitude of the torsion is related to the other two parameters of the rotation and this relation is described by Listing's and Donders's laws (Boring, 1942) [Bor41]. These laws imply that even if the torsion is not zero, its magnitude in the two eyes is very similar. Here, we assume that the magnitude of the torsion is the same in the two eyes. These two constraints allow us to remove two rotational parameters. Namely, for a given orientation of one eye, the orientation of the other eye is uniquely specified by the angle formed by the two visual axes (this angle is called vergence). Let the x-axis in each eye (camera) he parallel to the line connecting the two eyes. Then, the vergence angle (θ) is represented by the angle of the rotation of the object around the vertical (y) axis. Note also, that the position of one eye relative to the other eye is constant. Specifically, one eye is translated relative to the other eye along

a line parallel to the x-axis. This allows removing one additional parameter, namely, we can assume that the vertical translation T_y is zero.

We apply the above constraints to equation 2.3 to yield the following equation:

$$\begin{bmatrix} r_x & r_y & 1 \end{bmatrix} \begin{bmatrix} 0 & -T_z & 0 \\ T_z & 0 & -T_x \\ 0 & T_x & 0 \end{bmatrix} \begin{bmatrix} a & 0 & b \\ 0 & 1 & 0 \\ -b & 0 & a \end{bmatrix} \begin{bmatrix} l_x \\ l_y \\ 1 \end{bmatrix} = 0 \qquad (2.4)$$

Where a $= cos\theta$ and b $= sin\theta$.

2.3 Estimation of the Essential Matrix

We first rewrite equation 2.4 in the following way:

$$\begin{bmatrix} r_x & r_y & 1 \end{bmatrix} \mathbf{E} \begin{bmatrix} l_x \\ l_y \\ 1 \end{bmatrix} = 0$$

where

$$\mathbf{E} = \begin{bmatrix} 0 & -T_z & 0 \\ T_z & 0 & -T_x \\ 0 & T_x & 0 \end{bmatrix} \begin{bmatrix} a & 0 & b \\ 0 & 1 & 0 \\ -b & 0 & a \end{bmatrix}$$

E is known as the essential matrix. We then scale the essential matrix by $1/T_x$ and substitute $t = T_z/T_x$ to yield the following equation:

$$\begin{bmatrix} r_x & r_y & 1 \end{bmatrix} \begin{bmatrix} 0 & -t & 0 \\ at+b & 0 & bt-a \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} l_x \\ l_y \\ 1 \end{bmatrix} = 0 \qquad (2.5)$$

Finally, we perform appropriate grouping to yield the following equation:

$$\begin{bmatrix} r_y l_x & r_y & -r_x l_y \end{bmatrix} \begin{bmatrix} at+b \\ bt-a \\ t \end{bmatrix} = -l_y \qquad (2.6)$$

Note that equation 2.6 is only for one pair of corresponding image points with three unknown parameters: at+b, bt-a, and t. If three pairs of corresponding image points are available, then three equations are obtained that can be represented in the following matrix form:

$$\begin{bmatrix} r_{y1}l_{x1} & r_{y1} & -r_{x1}l_{y1} \\ r_{y2}l_{x2} & r_{y2} & -r_{x2}l_{y2} \\ r_{y3}l_{x3} & r_{y3} & -r_{x3}l_{y3} \end{bmatrix} \begin{bmatrix} at+b \\ bt-a \\ t \end{bmatrix} = \begin{bmatrix} -l_{y1} \\ -l_{y2} \\ -l_{y3} \end{bmatrix} \tag{2.7}$$

We will use Ax = e to represent equation 2.7, where x is a vector of the three unknown parameters that are the elements of the essential matrix in equation 2.5. If matrix A is of full rank, then there exists a unique solution

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{e} \tag{2.8}$$

With more than three points provided, we use the least squares method. Namely, we find x that minimizes the norm of (Ax - e). This is a standard linear regression problem with a solution

$$\mathbf{x} = (\mathbf{A}^t\mathbf{A})^{-1}\mathbf{A}^t\mathbf{b} \tag{2.9}$$

Once the essential matrix is computed, the parameters a, b, and t (rotation and translation) can then be solved simultaneously.

Note that the smallness of the norm of (Ax-b) is a necessary condition for the fact that the two images have been produced by the same shape. According to Duda and Hart's (1973) [DH73] terminology, the smallness of this norm is a quasi-invariant. Unlike quasi-invariants that involve a single image of an object (and the shape of the object, as well) [PR92] [Wei93] [PLed], this quasi-invariant operates on two retinal images and does not use the shape of the object. As a result, this quasi-invariant allows reconstruction of a novel shape.

2.4 Reconstruction

After the rotation and translation matrices are computed, the object point ($R_x R_y$R,) can be reconstructed as follows.

Consider equation 2.1 after incorporating constraints specified in Section 2 2.2. We multiply the matrices and obtain the following equation:

$$\begin{bmatrix} L_x \\ L_y \\ L_z \end{bmatrix} = \begin{bmatrix} aR_x + aT_x - bR_z - bT_z \\ R_y \\ bR_x + bT_x + aR_z + gaT_z \end{bmatrix} \tag{2.10}$$

Matrix equation 2.10 represents three ordinary equations. We take the ratio of the lefthand sides and of the righthand sides of the second and the third equations of matrix equation 2.10:

$$l_y = L_y/L_z = R_y/(bR_x + bT_x + aR_z + aT_z) \tag{2.11}$$

Next we substitute R, $= r_x R_z$, R, $= r_y R_z$, and solve for R,:

$$R_z = (aT_z l_y + bT_x l_y)/(al_y + br_x l_y - r_y) \tag{2.12}$$

We divide the numerator and the denominator in equation 2.12 by T_x and substitute t $= T_z/T_x$:

$$R_z = [(atl_y + bl_y)/(al_y + br_x l_y - r_y)]T_x \tag{2.13}$$

The remaining coordinates R_x and R, are found from:

$$R_x = r_x R_z \tag{2.14}$$

$$R_y = r_y R_z \tag{2.15}$$

Note that the coordinates of the object point expressed by equations 2.13, 2.14 and 2.15 are known up to a multiplicative factor T,. This means that these equations determine the shape of the object, but not its size and not its distance from the observer (i.e. depth).

3. SIMULATIONS

The new algorithm was tested in the task of reconstructing **3-D** objects. First, we describe the stimuli and the testing method, then we will present the results.

3.1 Stimuli

Twenty-five cylinders of revolution were used as stimuli. Each cylinder was generitted by rotating a distinct **2-D** cubic B-Spline curve about its vertical axis. The height of all objects was the same. Fig. **3.1** shows an example of the stimuli.

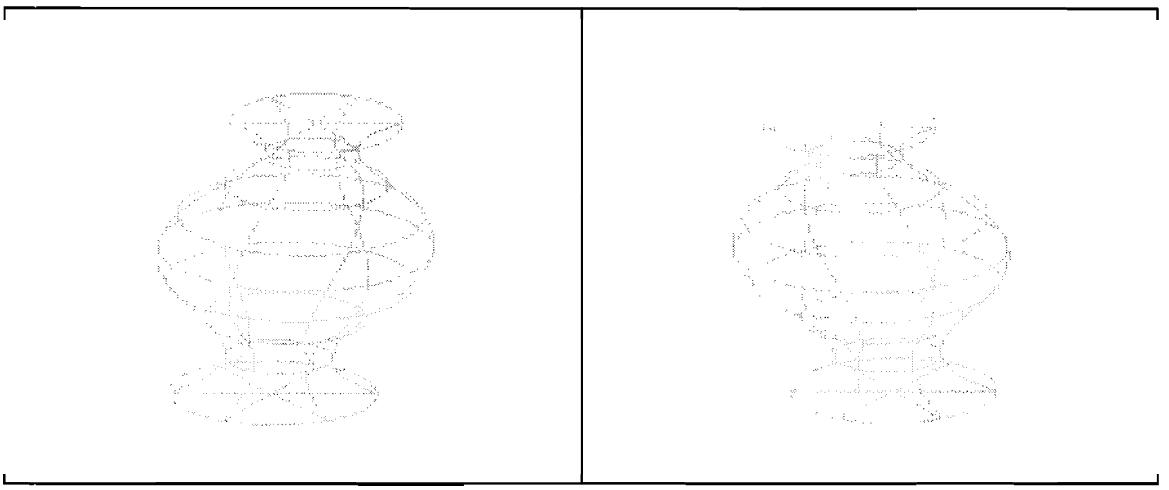

Fig. **3.1.** A sample stimulus represented by wireframe.

The camera images were computed according to the model shown in Fig. **2.** The simulated distance was approximately four times the object's size and the angle between the visual axes of the two cameras was 8 deg. These viewing parameters were identical to those used in the psychophysical experiment (see Section 4). Reconstruction was based on a number of points selected on the surface of each object. After the images of these points were computed, Gaussian noise was added to each image

coordinate with mean value zero and standard deviation proportional to the size of the object's image.

3.2 Testings

The testing involved 1000 trials for each object (total 25000 trials) for each condition. On each trial the object to be reconstructed (comparison object) was obtained from the original object by changing its aspect ratio. Specifically, the object was stretched or compressed along the direction of the axis of revolution and along the directions orthogonal to this axis. These transformations involved a scaling factor from within a range of 0.5 to 2.0. As a result of this transformation the comparison object was still a cylinder of revolution but with a different aspect ratio as compared to the original object. These scaling factors were randomly generated from trial to trial. After the object had been transformed, it was slanted so that its axis of revolution formed an angle of 45, 55 or 65 degrees with the frontal plane.

The accuracy of the reconstruction was evaluated by comparing the aspect ratio of the reconstructed object to that of the comparison object. In each trial, a ratio of the two aspect ratios was computed. The mean and the standard deviation of the mean were calculated at the end of each experiment.

3.3 Results

The results are shown in figures 3.2 - 3.5. The ordinate shows the average ratio of the two aspect ratios. The "one" on the ordinate represents an accurate reconstruction. The abscissa shows slant. Each data point is a mean computed from 1000 reconstructions. The standard deviation of the mean ratio in all experiments was quite small (less than 0.001). Thererfore, these standard deviations are not shown on the graphs.

Figures 3.2 and 3.3 show the algorithm's performance under different amounts of noise. The performance systematically deteriorates as the noise increases. Moreover, the performance is more stable when more points are used.

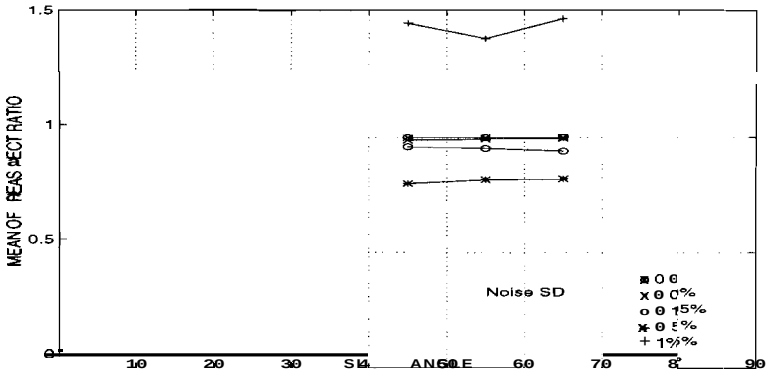

Fig. 3.2. Effect of noise on binocular reconstruction - 4 points

Figure 3.4 shows the effect of the viewing distance on shape reconstruction. Although the viewing distance is not an essential parameter in shape reconstruction when there is no noise in the image, the viewing distance affects the accuracy of shape reconstruction in the presence of noise. When the distance of the object from the camera. is large, then the vergence angle becomes so small that the differences between the left and right images produced by the difference in the viewing directions are small as compared to differences produced by noise. As a result, shape reconstruction deteriorates.

Finally, consider the effect of the object's size (Figure 3.5). If the object's size is small relative to the viewing distance, then perspective projection becomes approximately equivalent to affine transformation. Again, in the absence of noise in the images, this fact has no effect on the accuracy of reconstruction. However, if noise is present, it overshadows the perspective effects. Since, as pointecl. out in Section 1, two affine (or orthographic) images are not sufficient for unique reconstruction of an object, the reconstruction in the case of small objects when image noise is present, should be less accurate. This fact is shown in Figure 3.5.

In the next section we will present a psychophysical experiment which tested the psychological plausibility of our new model.

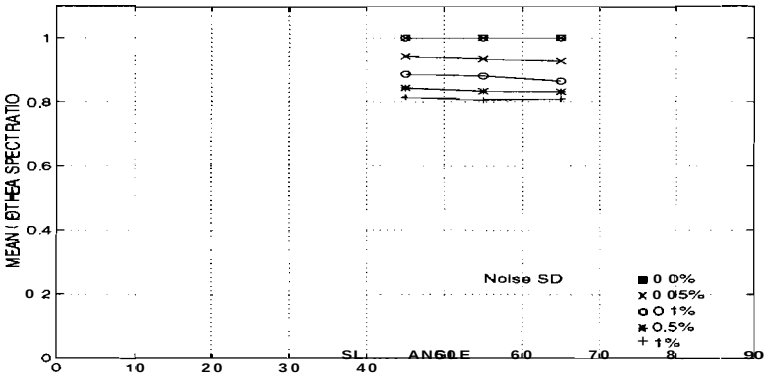

Fig. **3.3.** Effect of noise on binocular reconstruction - 9 points

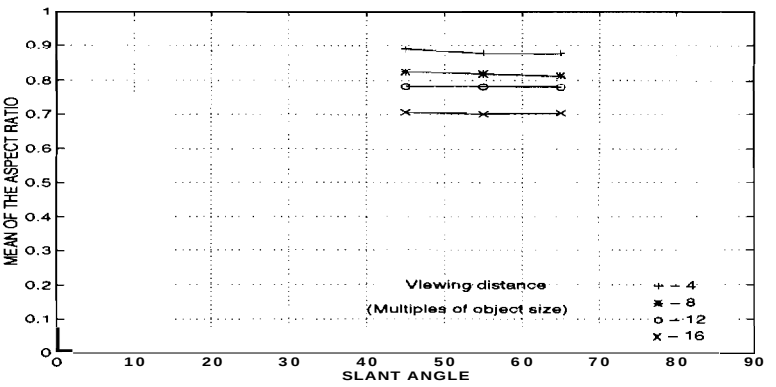

Fig. 3.4. Effect of viewing distance on binocular reconstruction - 9 points with noise standard deviation of 0.1 percent.

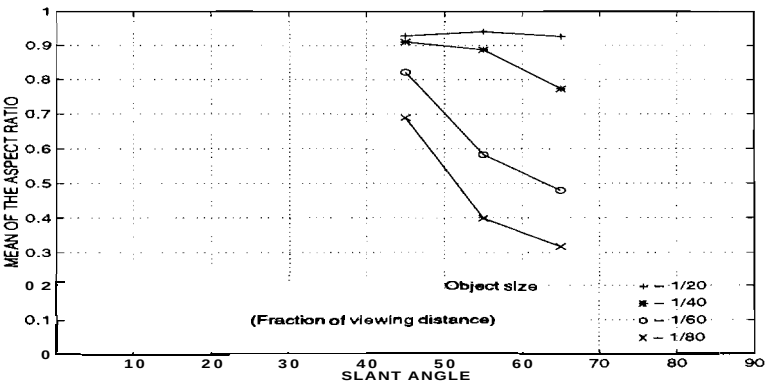

Fig. *3.5.* Effect of object size on binocular reconstruction - 9 points with noise standard deviation of 0.1 percent.

4. PSYCHOPHYSICAL EXPERIMENT ON SHAPE CONSTANCY

This experiment was performed in order to verify the psychological plausibility of our new algorithm. The main feature of our algorithm is that it reconstructs the shape of a 3-D object from two perspective images without using depth cues. This feature was tested in this experiment.

4.1 Method

4.1.1 Subject

The authors served as subjects in this experiment. They all had normal, or corrected to normal vision.

4.1.2 Stimuli

Cylinders of revolution, that were described in Section 3 were used as stimuli. The stimuli were displayed on a computer monitor using either monoscopic or stereoscopic (binocular) mode of viewing. The only difference between these two viewing modes was that no horizontal disparity was used in the monoscopic sessions. Three types of monocular cues were used: occluding contour, wireframe, and shading. Examples of these stimuli are shown in Figs. 3.1, 4.1 and 4.2.

If shape constancy involves reconstruction of the shape from depth cues, then the performance with wireframe and shaded stimuli should be much better than that with occluding contour. If, on the other hand, shape constancy involves quasi-invariants computed from distinctive points (like the quasi-invariant formulated in Section 2), then the performance with wireframe stimuli should be the best.

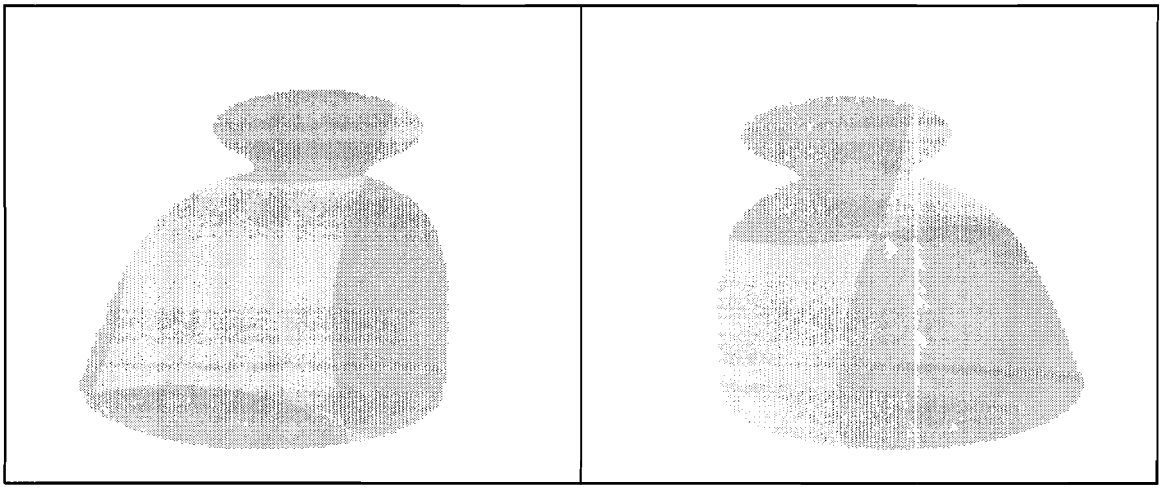

Fig. 4.1. A sample stimulus represented by occluding contour.

4.1.3 Viewing Conditions and Apparatus

The CrystalEyes system (StereoGraphics Corporation) [Lip91] was used for displaying stereoscopic images. This system consists of a pair of liquid-crystal display (LCD) glasses, a monitor, an infra-red emitter and a graphics display controller. Each LCD lens was electrically controlled to be opaque or transparent in synchronization with the display. The switching rate was 144/sec.

The luminance of the object through the active LCD glasses was 8.5cd/m^2, and the luminance of the background was 32.0cd/m^2.

4.1.4 Procedure

There were a total of six different sessions in this experiment representing all combinations of the viewing condition (monoscopic vs stereoscopic) and the monocular cue (occluding contour, wireframe and shading).

The order of the sessions was random and different for each subject. In each trial, the subject was presented with two objects. The standard object was displayed on the left with an upright orientation. The comparison object that was displayed on the right, was stretched and slanted the same way as described in Section 3. The subject's task was to adjust the height of the standard object to match the aspect

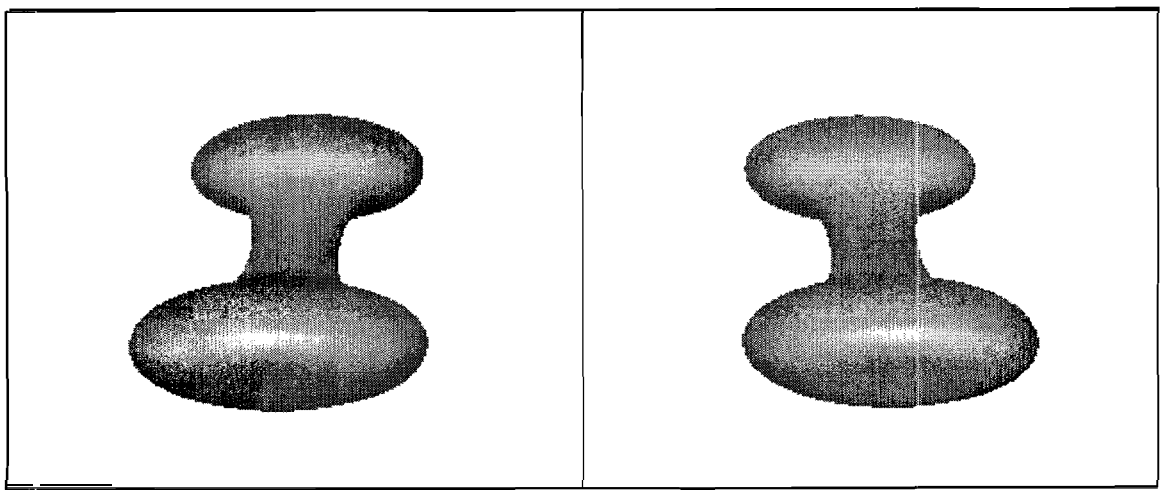

Fig. 4.2. A sample stimulus represented by shaded surface.

ratio of the comparison object. A ratio of the two aspect ratios was computed in each trial. This ratio was used as a measure of accuracy of the percept.

Each session contained a total of 75 trials. In each session, 25 different objects were displayed with different slant angles in a random order. Each object was displayed exactly three times and the slant angles were uniformly distributed from 40 to 70 degrees. The results from the 75 trials were grouped into three sets corresponding to the magnitude of slant: 40-50 deg, 50-60 deg, and 60-70 deg. The average ratio and the standard deviation of this ratio were computed from these sets.

The viewing distance was 50cm and the size of the simulated object was about 10 x 5 x $5cm^3$. The subject's head was supported by a chin-forehead rest.

4.2 Results and Discussion

Figures 4.3 - 4.8 show the. results. The ordinate shows the average ratio of the two aspect ratios. Thus, "one" on the ordinate represents an accurate percept. The abscissa shows slant. Each data point is a mean computed from 25 adjustments. The height of a symbol representing a mean judgment is equal to one standard deviation of the mean.

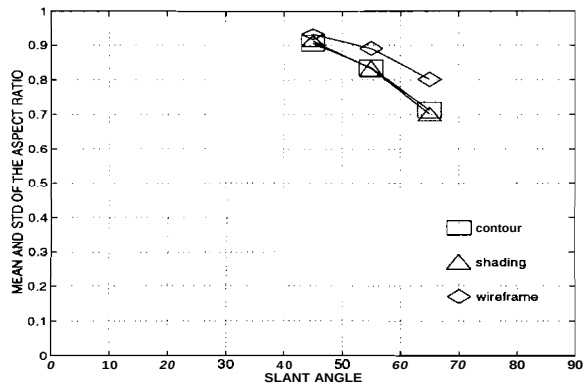

Fig. 4.3. MWC Monoscopic Viewing.

There are several results observed for all subjects. First, the performance deteriorated as the slant angle increased. Second, stereoscopic viewing led to more accurate and more constant percept as compared to monoscopic viewing (perfect shape constancy would be represented by horizontal lines). These two results are not new. What is new is that the performance was about the same regardless of whether the stimuli were represented by wireframe. shading (plus occluding contour) or just the occluding contour itself. This result is interesting because it is known that shading and wireframe are potentially very useful depth cues that can improve shape reconstruction. However, these cues do not appear to be important in human 3-D shape perception. This result is consistent with our algorithm because our algorithm performs shape reconstruction without using depth cues. Note also, that our psychophysical results suggest that human binocular shape constancy operates on occluding contours of the 3D objects, rather than on distinctive points. This fact cannot be explained by any of the existing theories of human or computer vision (including our new algorithm) and poses a challenge for the researchers of computational vision.

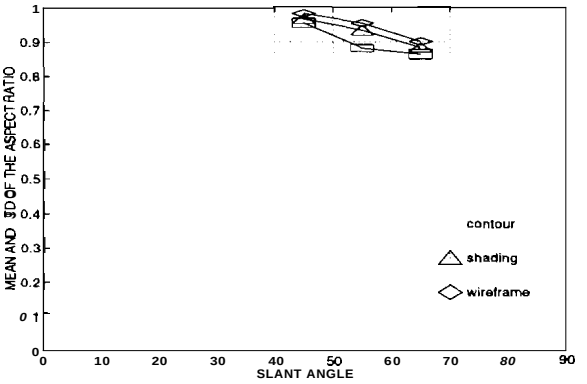

Fig. 4.4. MWC Stereoscopic Viewing.

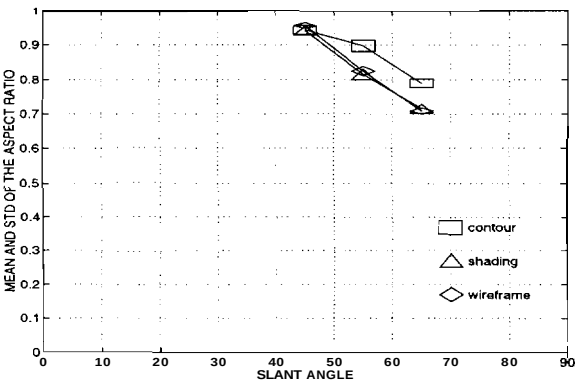

Fig. 4.5. ZP Monoscopic Viewing.

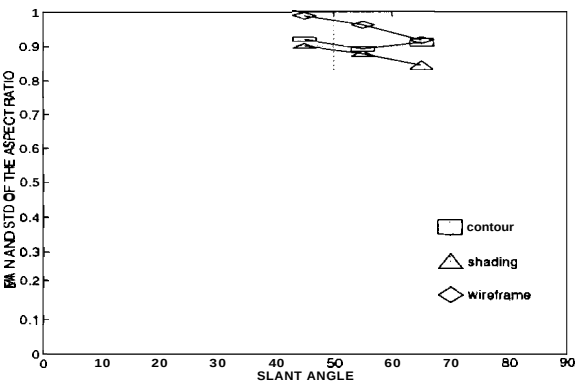

Fig. 4.6. ZP Stereoscopic Viewing.

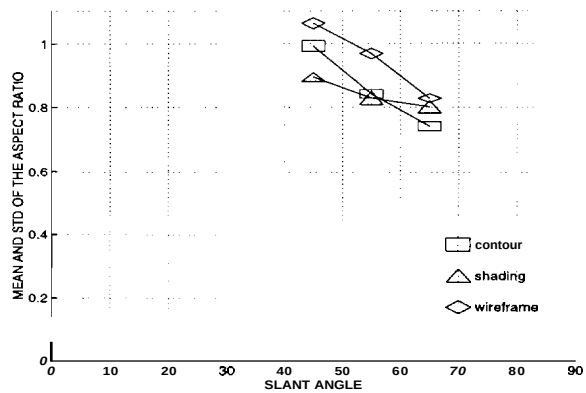

Fig. 4.7. DMC Monoscopic Viewing.

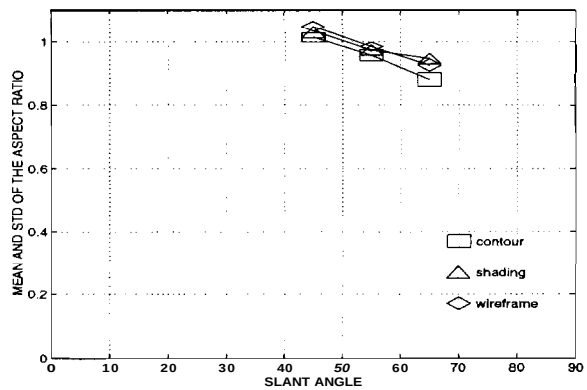

Fig. 4.8. DMC Stereoscopic Viewing.

5. SUMMARY

We reviewed prior research on shape recognition and reconstruction in both psychological and computational literatures. The psychological literature on shape constancy shows that depth cues are not involved in shape perception. Instead, shape constancy is more likely to involve quasi-invariants computed from a shape and its image. In this paper, we also used a quasi-invariant but this new quasi-invariant is computed from two perspective images. As a result, it allows reconstruction of the shape without using depth cues. Our algorithm incorporates natural constraints of the human visual system related to the fact that the system uses binocular fixation. The new algorithm requires as few as three points and performs reliably in the presence of noise.

We compared the performance of our algorithm to that of the human observers who were tested in the psychophysical experiment. The results of this experiment show that depth cues are not used either in monoscopic or stereoscopic viewing. This result is consistent with our new algorithm. Furthermore, the results show that human observers rely on the information derived from occluding contours of the object. This result is not explained by our algorithm and it will be investigated in our future research.

LIST OF REFERENCES

[AB89] J. Aloimonos and C. M. Brown. On the kinetic depth effect. Biological Cybernetics, 60:445–455, 1989.

[Ast95] Kalle Astrom. Fundamental limitations on projective invariant of planar curves. IEEE Transactions on Pattern Analysis and *Machine* Intelligence, 17-1:77–81, 1995.

[BBHP92] E. B. Barrett, E. H Brill, N. N. Haag, and P. M. Payton. Geometric invariance in *computer* vision by J. L. *Mundy* and A. Zisserman, chapter Invariant linear methods in photogrammetry and model-matching. MIT Press, 1992.

[BBP92] E. H. Brill, E. B. Barrett, and P. M. Payton. Geometric invariance in computer vision by *J.* L. *Mundy* and A. Zisserman, chapter Projective invariants in two and three dimensions. MIT Press, 1992.

[Bor41] E. G. Boring. Sensation and Perception in the History of *Experimental* Psychology. New York: Appleton, 1941.

[BWR90] Brian J. Burns, Richard Weiss, and Edward M. Riseman. View variation of point set and line segment features. In DARPA *Image* Understanding Workshop, pages 650–659, 1990.

[Cas44] E. Cassirer. The concept of group and the theory of perception. Philosophy and *Phenomenological* Research, 5:1–35, 1938/1944.

[CR41] R. Courant and H. Robbins. *What* is Mathematics? Oxford Press, 1941.

[Cut86] J. E. Cutting. Perception with an eye for *motion.* MIT Press, 1986.

[DH73] R. O. Duda and P. E. Hart. Pattern Classification and Scence Analysis, chapter 11. Wiley, 1973.

[Gib50] J. J. Gibson. The Perception of Visual World. Houghton Mifflin, 1950.

[Har95] Richard I. Hartley. In defence of the 8-point algorithm. In ICCV, pages 1064–1070, 1995.

[Joh77] G. Johansson. *Spatial constancy and motion in visual perception. In Epstein, W. (Ed), Stability and Constancy in Visual Perception*, chapter Mechanisms and Processes. Wiley, 1977.

[Joh91] E. B. Johnston. Systematic distortions of shape from stereopsis. *Vision Research,* 31-7/8:1351–1360, 1991.

[Kof35] K. Koffka. *Principles of Gestalt Psychology.* Harcourt Brace, 1935.

[KvD91] J. J. Koenderink and A. J. van Doorn. Affine structure from motion. *Journal of the Optical Society of America A*, 8-2:377–385, 1991.

[LH81] H. Christopher Longuet-Higgins. A computer algorithm for reconstructing a scence from two projections. *Nature,* 293-10:133–135, 1981.

[Lip91] Lenny Lipton. *The CrystalEyes Handbook.* StereoGraphics Corporation, 1991.

[Piz94] Zygmunt Pizlo. A theory of shape constancy based on perspective invariants. *Vision Research,* 34-12:1637–1658, 1994.

[PLed] Zygmunt Pizlo and Kirk Loubier. Recognition of a solid shape from its single perspective image using a quasi-invariant. Submitted.

[PR92] Zygmunt Pizlo and Azriel Rosenfeld. Recognition of planar shapes from perspective images using contour-based invariants. *Computer Vision; Graphics and Image Processing: Image Understanding*, 56:330–350, 1992.

[RFZM93] C. A. Rothwell, D. A. Forsyth, A. Zisserman, and J. L. Mundy. Extracting projective structure from single perspective views of 3d point sets. In *IEEE International Conference and Computer Vision,* pages 573–582, 1993.

[Roc83] I. Rock. *The logic of perception.* MIT Press, 1983.

[SA89] Minas E. Spetsakis and John Aloimonos. Optimal motion estimation. In *Proceedings of Workshop for Visual Motion,* pages 229--237, 1989.

[Sta45] B. K. Stavrianos. The relation of shape perception to explicit judgments of inclination. *Archives of Psychology,* 296:1–94, 1945.

[TH84] Roger Y. Tsai and Thomas S. Huang. Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 6-1:13–27, 1984.

[Tho59] E. H. Thompson. A rational algebraic formulation of the problem of relative orientation. *Photogrammetric Record,* 3-14:152–159, 1959.

[TK90] Carlo Tomasi and Takeo Kanade. Shape and motion without depth. In *International Conference of Computer Vision,* pages 258–270, 1990.

[TN94] J.T. Todd and J. F. Norman. The visual perception of 3d length in natural vision. *Investigative Ophthalmology and Visual Science.,* 34:1131, 1994.

[Ull79] S. Ullman. *The Interpretation of Visual Motion.* MIT Press, 1979.

[Wei88] Issac Weiss. Projective invariants of shapes. In *IEEE Conference on Computer Vision and Pattern Recognition,* pages 291–297, 1988.

[Wei93] D. Weinshall. Model-based invariants for 3-d vision. *International Joural of Computer Vision,* 10:27–42, 1993.